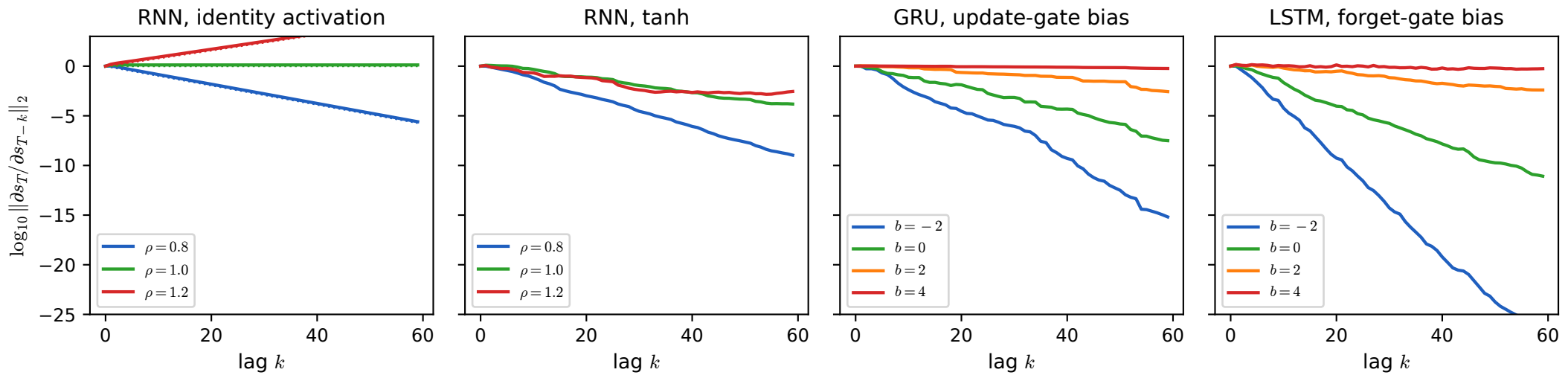

RNN, identity activation
RNN, tanh
GRU, update-gate bias
LSTM, forget-gate bias
$\log_{10} \|\partial s_T / \partial s_{T-k}\|_2$
0
−5
−10
−15
−20
−25
0
20
40
60
lag k
$\rho = 0.8$
$\rho = 1.0$
$\rho = 1.2$
$b = -2$
$b = 0$
$b = 2$
$b = 4$